

Valsts drošības komitejas zinātniskās izpētes komisija

WWW.LU.LV/VDKKOMISIJA

ZIŅAS

Sākums > Ziņas

Ilmārs Poikāns: Ēsi par LPSR VDK dokumentu digitalizāciju

LPSR VDK zinātniskās izpētes komisija
07.02.2017

Latvijas Universitātes Matemātikas un informātikas institūta pētnieks, LPSR VDK zinātniskās izpētes komisijas loceklis M. sc. comp. Ilmārs Poikāns ir sagatavojis konspektīvu informāciju par LPSR VDK dokumentu un citu PSRS okupācijas varas represiju un institūciju dokumentu digitalizācijas problēmjautājumiem.

Apskats *.pdf datnes formā pieejams šeit. Plašāka informācija Kristīne Jarinovska +3712977788, Ilmārs Poikāns +37129152972.

Lielākais LPSR Valsts drošības komitejas un citu PSRS okupācijas represīvo iestāžu dokumentārā mantojuma apjoms glabājas Latvijas Nacionālā arhīva struktūrvienībā Latvijas Valsts arhīvā — dažādos fondos atrodas vairāk nekā 150 tūkstoši lietu un vairāk nekā 1 miljons kartiņu. Materiāli par LPSR Valsts drošības komitejas darbību atrodas arī Personāla Dokumentu Valsts arhīvā, reģionālajos arhīvos un citur.

Digitalizācija iekļauj visas darbības, kas saistītas ar digitālās bibliotēkas veidošanu — materiālu pārvēršanu ciparu formā, metadatu pievienošanu, pieejamības nodrošināšanu, ilgtermiņa saglabāšanu.

Kāpēc jādigitalizē?

Digitalizēti materiāli ir daudz pieejamāki pētniekiem un interesentiem, jo tiem var piekļūt no jebkura datora ar interneta pieslēgumu — tiek ietaupīts gan laiks, gan nauda, jo nav jābrauc uz materiālu glabātāvām. Taupīts tiek arī Latvijas Nacionālā arhīva cilvēkresurs, Latvijas Universitātes un citu zinātnisko institūciju vēsturnieku cilvēkresurs — gan arhivisti, gan vēsturnieki var izmantot savu intelektuālo potenciālu savam pamatdarbam, nevis savu laiku patērēt primitīvām un tehniskām darbībām. Turklāt, ne visiem fiziskiem materiāliem tiek nodrošināta pieeja, pat esot uz vietas arhīvā.

Materiālu digitalizācija palīdz saglabāt oriģinālus, kas var tikt sabojāti un pat neatgriezeniski zaudēti gan fiziska nolietojuma rezultātā, antropomorfās ietekmes dēļ, gan laika zoba — izdziedošas tintes, papīra sairšanas, ķīmiskā zīmju dzišanas, fotogrāfiju balēšanas — dēļ.

Gadījumā, ja kaut kādu iemeslu dēļ iet bojā oriģināli, tad vismaz vēl ir pieejama digitāla kopija, kurā var iegūt to pašu informāciju. Šajā gadījumā īpaši svarīga ir veiktās digitalizācijas kvalitāte un pilnīgums.

Digitālus materiālus ir iespējams ievietot vienotā informācijas krātuvē, kas ļauj ērtāk un pilnvērtīgāk veikt pētījumus. Šāda sistēma var arī atvieglot pētniecības gaitā paveiktā darba (teksta atšifrēšana, vietu un personu atpazīšana u.c.) saglabāšanu un atkārtotu izmantošanu citos pētījumos. Ar metadatiem papildināti materiāli ļauj vieglāk pamanīt un atrast sakarības starp dažādiem avotiem. Savukārt materiāli, ir sevišķi kartotēkas, kas ir pārvērstas strukturētā formā, kas atbalsta dažādu datu prasījumu veikšanu, ļauj veikt dziļākus kvantitatīvos pētījumus.

Kas jādigitalizē?

Visu nav iespējams digitalizēt uzreiz, tāpēc jāizvēlas materiāli, kas visvairāk atvieglos un paātrinās pētnieku darbu — kartotēkas, konkrētu fondu aprakstus vai pēc noteiktiem kritērijiem atlasītas arhīva lietu kopas.

Kartotēkas ir nozīmīgas kā primāra uzziņu sistēma pētniecības procesā. Tāpēc, digitalizējot materiālus, prioritāte dodama tieši kartotēku digitalizācijai. Tas ir arī salīdzinoši vienkāršāk veicams darbs, un to vēl ir iespējams sadalīt apakšposmos.

Priekšroka dodama arī ātri digitalizējamām lietām, piemēram, dokumentiem uz atsevišķām lapām vai korekti iesietām lietām, kuras nav jāizjauc, lai tās varētu ieskenēt līdzīgi kā grāmatas.

Katram pētniekam būtu jāzina, kādi materiāli viņa pētījumos būtu noderīgi un kāda ir to savstarpējā prioritāte no digitalizācijas viedokļa. Apkopojot visu pētnieku vajadzības un vēlnes, var izveidot precīzāku un digitalizācijas plānu ar vislielāko atdevi.

Pielikumā 1. tabulā ir saraksts ar LNA Latvijas Valsts arhīvā esošiem VDK, leTK izcelsmes un ar tiem saistītiem LKP dokumentu fondiem.

Kā jādigitalizē?

Materiālu digitalizācijas rezultātā jāiegūst ticamas digitālās reprodukcijas, kas pēc iespējas precīzāk atveidotu oriģinālo objektu gan pēc satura, gan arī pēc izskata. Svarīga ir objektu un to sastāvdaļu savstarpējās saistības un vietas saglabāšana, lai jebkuru materiālu var atainot tā, kā tas ir bijis oriģinālā. Jo precīzāk atveidots oriģināls, jo lielāka iespēja, ka pētniekam pietiks ar digitālo kopiju, lai veiktu specifisku objektu analīzi. Arhīvdarbes ir paredzētas ilgtermiņa glabāšanai un atvasināto kopiju veidošanai. Tās parasti ir nesaspīestas un maz vai nemaz rediģētas pēc objekta digitalizēšanas. Lietotājdarbes tiek iegūtas no arhīvdarbnēm un ir paredzētas publicēšanai tīmeklī un aplūkošanai uz ekrāna.

Jānodrošina iegūto digitālo materiālu ilgtermiņa saglabāšana (desmitiem un simtiem gadu). Vēlamas vismaz trīs kopijas uz diviem dažādu veidu datu nesējiem, un viena kopija jāglabā citā vietā. Jārēķinās arī ar datu krātuves uzturēšanas izmaksām — neizbēgama migrācija uz jauniem datu nesējiem, kad esošie būs nokalpojuši savu mūžu. Izmantojot plaši pazīstamus un atvērtus datu formātus kā TIFF (attēliem), XML (semantiskai informācijai), būs arī vieglāka migrācija uz jaunām sistēmas komponentēm, kas parasti uztur importu un eksportu šādos formātos.

Viens no veidiem, kā nodrošināt materiālu satura ilgmūžību, neatkarīgi no iespējamām tehniskām kļūdām, datu nesēju un failu formātu problēmām, ir izveidot papildus kopijas uz mikrofilmām. Tā kā ilgtermiņa glabāšanai piemērotākas ir melnbaltās mikrofilmas, bet ir dokumenti, kurus jādigitalizē krāsās, tad būtu ieteicams mikrofilmas veidot no digitalizācijas rezultātā iegūtiem attēliem, nevis veikt digitalizāciju, uzņemot un vēlāk skenējot mikrofilmas. Pēdējai pieejai ir vēl viens trūkums — tā pēc ātruma būtiski zaudē dokumentu skeneriem, ar kuriem efektīvi var skenēt kartotēkas.

Skenēšanas operatoriem jāvadās pēc iepriekš izstrādātas metodikas, lai ar minimālu piepūli būtu uzreiz skaidrs, ko digitalizēt, bet ko nē. Tas ir būtiski, lai dokumentus bez aizķeršanās varētu skenēt operatori, kuriem nav specifisku zināšanu par konkrēto dokumentu grupu. Darbības, kas prasa iedziļināšanos skenējamā materiālu saturā un kaut kāda lēmuma pieņemšanu, rada gan nevajadzīgu kognitīvo slodzi operatoram, gan paver iespēju kļūdām lēmuma pieņemšanā. Bet tā kā jebkura materiāla pārskenēšana ir būtiskas papildizmaksas un zaudēts laiks, tad izdevīgāk būs izdarīt visu pēc iespējas pilnīgāk ar pirmo reizi.

Ir obligāti jāskenē visas lapas, kas norādītas dokumenta aizmugurē esošajā apliecinājuma lapā. Jāskenē arī pati apliecinājuma lapa. Svarīgs ir jautājums, ko darīt ar nenumurētajām lapām, kas ir tukšas. Labāk ir ieskenēt arī tukšās lapas, lai vēlāk izslēgtu iespēju, ka kaut kādā veidā tiek apstrādāts ieskenētais dokuments, pamatojoties ar to, ka trūkst lapas (tās nav ieskenētas).



LPSR VDK zinātniskās izpētes komisijas loceklis M. sc. comp. Ilmārs Poikāns 2016. gada 7. decembrī, Kultūras ministrijā, komisijas rakstu 3. sējuma atvēršanā. Kristīnes Lindbergas foto.

Piemēram, skenējot kartotēkas, tiek automātiski ieskenēta arī kartītes otra puse, kurā dažos gadījumos nekā nav, taču to ieskenējot un glabājot kopā abas puses vienā failā ir izveidota viena pabeigta un neapstrīdama arhīva dokumenta vienība. Tātad — ja skenē, tad visu pēc iespējas pilnīgāk, jo iespējama ietaupījums kaut ko neieskenējot ir mazs, salīdzinot ar papildus izmaksām, ja nāksies vēlreiz atgriezties pie lietas materiāliem un skenēt kaut ko papildus, kas netika izdarīts pirmajā reizē.

Lai minimizētu iespēju, ka ar izjauktām lietām notiek nevēlami negadījumi, nepieciešams minimizēt laiku, kad lieta atrodas izjauktā veidā. Nepieciešama tāda darba plūsma, kurā lietu kā sāk digitalizēt, tā to arī pēc iespējas ātrāk pabeidz.

Esošo datu izmantošana

Daudziem pētniekiem ir dati no iepriekšējiem pētījumiem. Ja ir iespējams, tad tos noteikti ir jādara pieejamus pēc iespējas plašākam pētnieku lokam.

Daļā gadījumu šie dati tiek glabāti kādā no novecojušiem failu formātiem un tos nepieciešams migrēt uz jaunākiem, lai tos varētu pilnvērtīgi izmantot. Piemēram, ir veikta Kārļa Kangeras rīcībā esošās datubāzes, kurā glabājas grāmatas "No NKVD līdz KGB"[1] dati, migrācija no vecas Filemaker versijas uz mašīn- un cilvēkvasāmu mūsdienās plaši lietotu datu formātu (csv, xls).

Digitalizācijas posmi

Materiālu digitalizācija nav iespējama uzreiz pilnā apjomā. To ir jāsadala loģiskos posmos tā, lai pēc katra no posmiem būtu reāls ieguvums lietotājiem.

Kartotēku digitalizācija ir salīdzinoši vienkāršāka nekā citu veidu materiālu digitalizācija. Pirmais solis ir vienkārša kartotēkas kartīšu ieskenēšana, radot iespēju pētniekiem šķirstīt kartotēku digitālā veidā. Otrais solis ir kartīšu satura atpazīšana un glabāšana strukturētā formātā. Tas ļauj pētniekiem veikt kvantitatīvus pētījumus dažādos griezumos, kas bez digitalizēto materiālu datoranalīzes ir iespējami tikai izlases formā.

Runājot par visu materiālu digitalizāciju, tad pirmais posms ir materiālu ieskenēšana un attēlu iegūšana, lai būtu iespējama vienkārša šķirstīšana attālināti.

Otrais posms ir ieskenēto materiālu tekstu automātiska atpazīšana, ja tas vispār ir iespējams. Automātiski atpazīto tekstu ar visām kļūdām var izmantot, lai izveidotu vienkāršotu meklēšanu brīvā tekstā. Neskatoties uz atpazīšanas kļūdām, šāda iespēja meklēt brīvā tekstā noder pētniecības darbā. Atpazīšanas kļūdu negatīvo ietekmi var mazināt, meklēšanu veicot ar paplašinātu vaicājumu, kas pieļauj noteiktu skaitu kļūdu vārdos.

Pēdējais posms sastāv no vairākiem apakšposmiem, kurus var izpildīt jebkurā secībā atkarībā no pētnieku prioritātēm.

Pilnīgu metadatu pievienošana par objektiem noder, lai savienotu dažādus materiālus ar dažādām personām un vietām, kā arī notikumiem un laikiem. Piemēram, lietā ir informācija par personu, izmeklētājiem, apcietināšanas vietu un laiku, liecībās tiek pieminētas citas personas un notikumi u.t.t. Jo vairāk lietām un materiāliem ir pievienoti šādi metadati, jo interesantākus notikumus un sakarības iespējams atklāt un analizēt, izmantojot saistību grafu.

Automātiski atpazītā teksta labošana. Var izstrādāt rīkus, kas meklē tipiskās kļūdas un tās izlabo automātiski vai pēc manuāla apstiprinājuma. Taču ir kļūdas, kuras var atklāt tikai pārļausot tekstu. Tā kā uzreiz veikt visu tekstu pārbaudi ir neiespējami, tad praktiskākā pieeja ir ļaut lietotājiem izlabot viņu pamanītās kļūdas materiālos, ar kuru viņi dotajā brīdī strādā. Tas ir iteratīvs process, taču katra izlabotā kļūda ir lai arī mazs, tomēr uzlabojums. Un darbs dabiskā veidā tiek ieguldīts biežāk izmantotajos materiālos.

Būs gadījumi, kad svarīgus dokumentus nevarēs atpazīt automātiski. Tāpēc jābūt iespējai saturu ievadīt pilnībā manuāli.

Strukturēto datu pārvēršana datubāzes vai kādā citā formā, kas atbalsta vaicājumu izpildi. Pētniekiem bieži ir noderīga tāda informācijas sistēma, kura ir pieskaņota viņa konkrētajam vajadzībām konkrētajā pētījumā. Ja dati ir pieejami strukturētā veidā, tad tas paver iespēju izveidot pētniekam nepieciešamos skatus uz jau esošiem datiem.

Pilnvērtīga meklēšana pēc dažādiem parametriem. Ideālā gadījumā meklēšanas formu konfigurēšanai jābūt tik vienkāršai, lai parasts lietotājs var izveidot specializētu meklēšanas formu tieši viņa vajadzībām.

Pašreizējie darbi

Tiek skenēta Latvijas represēto iedzīvotāju kartotēka (LVA-1986). Šī ir vissarežģītākā kartotēka no sagatavošanas darbu un digitalizācijas viedokļa, jo satur fiziski noliegtas un dažāda izmēra kartītes, kā arī aplodes, kas prasa lielu manuālu pirms un pēcstrādi.

Notestēta LKP CK pilsētu partijas komiteju, LKJS fondu kartotēku skenēšana. No tehniskā viedokļa tās ir vieglāk skenējamas.

Praktiskā lielapjoma darbā ir veiksmīgi notestēti dokumentu skeneri Canon DR-G1100, Canon DR-6030C (ar tiešo caureju aplocēm un īpaši biežām kartītēm).

Notestēti grāmatu skeneri, drīz būs pieejams pēc gataviem rasējumiem izgatavots Archivist skeneris, kas skenē materiālus pa atvērumiem ar divām foto kamerām.

Tiek veikta vienotās informācijas sistēmas koncepcijas izstrāde.

Digitalizācijas šķēršļi un sarežģījumi

Finansējums

Pašreiz ar VDK saistīto materiālu digitalizācijai nav paredzēti līdzekļi ne VDK zinātniskās izpētes komisijas, ne Latvijas Valsts arhīva budžeta ietvaros. Operatīvs papildus finansējums tieši plašas digitalizācijas mērķiem, VDK zinātniskās izpētes komisijai ļautu līdz savas darbības beigām 2018. gadā ielikt stingrus pamatus efektīvākai un dziļākai pieejamo materiālu izpētei.

Juridiskie jautājumi

Lielāko daļu lietu, kas atrodas Latvijas Valsts arhīvā, bez izjaukšanas nav iespējams pilnvērtīgi digitalizēt. Nepieciešams izstrādāt skaidru juridisko regulējumu, kas tiek piemērots izjaucot un saliekot atpakaļ lietas, lai arhīva dokumenti pēc tam netiktu apstrīdēti no uzticamības un juridiskā viedokļa.

Uzglabāšanas vietas palielināšana

Izjaucot lietu, lai to pilnvērtīgi ieskenētu, nav iespējams to tik pat kompakti salikt atpakaļ, kāda tā bija sākumā. Lietu izjaukšana un salikšana atpakaļ palielinās arhīvā aizņemto vietu jeb "lineāros metrus". Tas prasīs ne tikai papildus telpas, bet arī esošo materiālu pārkārtošanu, jo glabātāvās dažādi fondi stāv secīgi viens aiz otra — ja viens fonds paliek garāks, tas automātiski rada kaskādes efektu, kā rezultātā var nākties pārkārtot tam sekojošos fondus.

Tehniskie sarežģījumi

Kartotēku digitalizācija ir jau notestēta — vissarežģītākā no digitalizācijas viedokļa ir Latvijas represēto iedzīvotāju kartotēka (LVA-1986), kurā lielā daudzumā tiek izmantotas aplodes (satur kopā vairākas kartītes par vienu un to pašu personu). Pārējās kartotēkās aplodes vai nu netiek izmantotas, vai arī tiek izmantotas daudz mazāk, tāpēc to digitalizācija būs vieglāka.

Iesietu lietu digitalizācija ir daudz sarežģītāka. Papildus tam, ka lietas būs nepieciešams izjaukt un salikt atpakaļ, ir konstatēti dažādi gadījumi, kas apgrūtina digitalizēšanu — puscaurspīdīgs papīrs, pielīmētas izziņas, salocītas lielāka izmēra lapas, bukleti, 3-dimensiju objekti, fotogrāfijas, filmas, vienas lapas saturs var daļēji turpināties uz blakus lapas, aizlocīti lapu sāni, daļēji aizsedzot saturu, kartes, milzīgi dokumenti, plakāti. Piemērus ar minētajiem piemēriem var redzēt pielikumā.

Cauršūts saturs

Lapas saturs daļēji aizsegts, jo tas atrodas pārāk tuvu iesūvumam, vai pat tas ir cauršūts. Šis ir viens no visbiežāk sastopamajiem sarežģījumiem. Ir tikai viens variants, kā iespējams pilnvērtīgi digitalizēt saturu šādos gadījumos — arhīva lieta ir jāizjauc un tad jāskenē atsevišķas lapas.

Puscaurspīdīgs papīrs

Lietās ir sastopamas iešūtas puscaurspīdīgas lapas, kuras rada sekojošu problēmu digitalizācijas procesā — uz šādu lapu skatoties, cauri var spīdēt ne tikai teksts no otras lapas puses, bet pat arī nākamo lapu saturs.

Nepieciešams likt tukšu baltu lapu aiz skenējamās lapas. Var likt arī melnu lapu, kas attēlu padarīs tumšāku, taču mazāk varēs redzēt tekstu otrā lapas pusē. Jāņem vērā, ka šāda papildus lapu likšana aiz skenējamām lapām ir laikietilpīgs process.

Problēma izpaužas spilgtāk, ja tiek skenēts spilgtā apgaismojumā, piemēram, uz planšetkenera. Vislabākais veids kā skenēt šādas lietas materiālus ir ar planetāro skeneri vai fotoaparātu.

Lapu "vēdekļi"

Mazākas lapiņas parasti ir iesietas "vēdekļa" formā, pakāpeniski nobīdot tās attiecībā vienai pret otru. Tas nozīmē, ka, skatoties uz šādu "vēdekli" atvērumā, var vienlaicīgi redzēt saturu no vairākām lapām. Lai ieskenētais materiāls būtu pēc iespējas vieglāk un ērtāk skatāms, šādos gadījumos skenējot katru no lapām "vēdekli", aiz tās jāliek balta lapa, lai izolētu skenējamās lapas saturu no pārējā.

Pielīmētas lapas un lapiņas

Sastopamas lietas, kurās ir mazāka formāta lapas vai lapiņas, piemēram, tālsarunu kvītis. Tās var būt tik mazas, ka tās nav iesietas kā lapu "vēdekļi", bet gan pa tiešo pielīmētas ar līmlenti vai līmi pie papīra lapas. Tās nav iespējams atraut no pamatlapas, to neobjādot. Šādas lapiņas var nebūt iespējams skenēt no augšas vienā piegājienā, ja šīs lapiņas pārkļājas

Ir gadījumi, kad ar pielīmētām lapiņām tiek aizsegta informācija uz pamatlapas vai pamatlapu nevar redzēt visu vienlaicīgi.

Aizlocītas malas un salocītas lapas

Šādām lapām parasti ir nodilušas malas, un katra iztaisnošana un salocīšana rada papildus bojājumus. Atlokot malas vai visu dokumentu, nepieciešams lapu iztaisnot — tā neturas līdzena. Viens no iespējamiem risinājumiem — izmantot plakanvirsmas skeneri vai fotografēt caur piespiedējistīku. Tas būtiski palēnina digitalizācijas procesu.

Anketas

Daļa arhīva lietu satur personas anketu, piemēram, filtrācijas lietas (LVA-1821), deportēto lietas (LVA-1894, LVA-1987, LVA-1994), par sevišķi bīstamiem pretvalstiskiem noziegumiem apsūdzēto personu krimināllietas (LVA-1986), nomenklatūras personas lietas (LVA-PA-15500). Šādās anketās daļa informācijas var nebūt redzama lapu iešūšanas dēļ. Skenēšana varētu būtu ātrāka ērtāka, ja varētu skenēt atvērumus un automātiski pareizi sanumurēt lapas. Jāņem vērā, ka ir A4 izmēra anketas, kuras atvērums ir A3 izmērā. Ir gadījumi, kad anketa ir aizpildīta tā, ka teksts tiek turpināts anketas atvēruma blakus lapā.

Mašīnraksts caur koppelāru

Sastopami teksti mašīnrakstā, kuros teksts ir ļoti izplūdis — tas ir viens no eksemplāriem, kas iegūts drukājot uz vairākiem koppelāru un papīra slāņiem. Šādos gadījumos automātiskā teksta atpazīšana nestrādā vai strādā ārkārtīgi slikti.

Lai arī mūsdienu OCR programmas atbalsta teksta atpazīšanu mašīnrakstā, taču tam jābūt skaidram. Neskaidri un izplūduši krievu valodas burti vismaz ABBY Finereader 11 programma netika atpazīti korekti — kļūdu skaits bija tik milzīgs, ka rezultāts praktiski nebija izmantojams nemaz. Papildus trenēšana uz izplūdušiem tekstiem iespējams varētu uzlabot precizitāti.

Otrs no iespējamiem variantiem, kā uzlabot šādus attēlus, lai tos varētu veiksmīgāk atpazīt, ir izmantot izplūdušu un defokusētu attēlu atjaunošanas algoritmus[2].

Dziestošā tinte

Sastopami dokumenti, kuri rakstīti izmantojot ķīmisko zīmuli vai, acīmredzot, taupības nolūkos atšķaidītu tinti. Tāds raksts ir izbaļājis un dažreiz ir ļoti grūti salasāms. Līdzīga problēma ir arī ar pirmajām 1950-to gadu lodišu pildspalvām. Izdzīst arī tas, kas rakstīts ar flomāsteriem. Situācija ar laiku kļūst arvien problemātiskāka, līdz teksts praktiski būs izdzisis. Tāpēc vēlreiz jāuzsver digitalizācijas nozīmīgums ne tikai pētnieciskā darba uzlabošanai, bet tieši vēsturiskā mantojuma saglabāšanai.

Personas dokumenti

Arhīva lietās ir atrodamas pases, karavīra dienesta apliecības u.c. personas dokumenti. Ir sastopami gadījumi, kad šādi dokumenti ir nevis ielikti kādā aploksnē, no kuras tos ir iespējams izņemt ārā, bet gan iešūti lietā.

Aploknes

Lai pilnībā digitalizētu arhīva lietas, no katras aploksnēs jāizņem viss tās saturs, un, saglabājot pareizo secību, jādigitalizē. Tas būtiski var palēnināt digitalizācijas ātrumu.

Lietiskie pierādījumi

Lietas var saturēt visdažādākos lietiskos pierādījumus. Ir lietas, kurās tie nāk līdz pat atsevišķās kastēs.

Kinolentes, foto filmiņas un foto

Lai digitalizētu kinolentes, foto filmiņas, nepieciešams atsevišķs digitalizēšanas process, iekārtas un palīg līdzekļi. Ņemot vērā, ka nepieciešamā infrastruktūra šādu darbu veikšanai var atrasties atsevišķā vietā, pat citā mājā, jāspēj organizēt vai nu drošs materiālu transportēšanas veids un darba plūsmas process, vai arī nepieciešamās iekārtas uz laiku jāatved un jāuzstāda tur, kur notiek materiālu digitalizācija.

Latvijas Nacionālā arhīva dokumentu preventīvās saglabāšanas departamenta mikrofotokopēšanas nodaļai ir laba pieredze un piemērotas iekārtas fotofilmu digitalizācijā. Kinofotofonodokumentu arhīvā notiek filmu digitalizācija. Lielākā problēma varētu būt ar specifisku kinolenšu veidu digitalizāciju, jo eksistē daudz un dažādi veci kinolenšu veidi.

Lielformāta izdevumi

Lietu materiālos un pierādījumos var atrast arī plakātus, avīzes u.c. lielformāta izdevumus. Plakanvirsmas skeneri līdz A3 izmēram ir pietiekami plaši pieejami, taču lielformāta plakanvirsmas skeneri (A2-A0 formāts) ir daudz dārgāki. Var uzskatīt, ka tie būs specializētas iekārtas, kuras lietošanā būs jārisina tie paši jautājumi kā kinolenšu un foto filmiņu skenēšanā.

Lielformāta izdevumus iespējams arī fotografēt, ja nav nepieciešama liela (150+ dpi) izšķirtspēja.

Vienotā informācijas sistēma

Salasot informācijas druskas, kas iegūtas no dažādiem materiāliem, avotiem un analīzes rezultātā, var iegūt daudz pilnvērtīgāku priekšstatu par pētāmo tēmu. Lai to izdarītu, nepieciešama vienotā informācijas sistēma.

Digitalizācijas un dažādu pētīnieku (ne tikai VDK zinātniskās izpētes ietvaros, bet arī neatkarīgu) darba procesā iegūtos un radītos metadatus un papildinātus nepieciešams saglabāt vienotā informācijas sistēmā. Tas nozīmē, ka praktiski jebkuras pētīnieka darbības, ko viņš veic sava pētījuma ietvaros, rezultāts tiek saglabāts. Piemēram, pētot vairākas filtrācijas lietas, pētīnieks noskaidro apcietināšanas vietu, šo informāciju jāsavstarpina vienotajā informācijas sistēmā.

Pētīniekiem, strādājot ar materiāliem un tos pētot, bieži ir noderīga kāda informācijas sistēma, kura ir pieskaņota viņa konkrētajam vajadzībām. No vienotās informācijas sistēmas varētu ērti veidot nepieciešamos skatus uz datiem, lai pētīnieks

ievadītu ne tikai datus, bet arī iegūtu nepārprotamu labumu savam darbam. Tad arī pētnieks to, kas nepieciešas kopējai lietai, darītu nevis kā papildus darbu, bet gan ar motivāciju, ka tas viņam pašam palīdzēs, jo sistēma spēs pētnieka ievadītos datus tam pašam pētniekam vēlāk pasniegt viņam ērtā formā.

Īpaši būtiskas ir sakarības starp dažādiem subjektiem, objektiem, vietām, notikumiem un laikiem dažādos materiālos. Šādas sakarības visērtāk ir praktiskāk uzglabāt grafu datubāzē, kam jābūt sistēmas pamatā.

Grafu datubāze ir datubāze, kas izmanto grafu struktūras semantiskiem vaicājumiem attiecībā pret virsotnēm (nodes), šķautnēm (edges) un īpašībām (properties), ar kuru palīdzību tiek reprezentēti un glabāti dati. Grafu datubāzes parasti ir viegli paplašināt ar iepriekš neparedzētiem laukiem un saītēm starp objektiem.

No šādas grafu datubāzes nepieciešamības gadījumā var eksportēt nepieciešamo informāciju formā, kas piemērota relāciju datubāzēm. Viegļāk būs izveidot sava veida skatus uz grafu datubāzi, nekā uzturēt vairākas relāciju datubāzes un tad tās datus sinhronizēt uz vienu grafu datubāzi.

Vienotajai informācijas sistēmai jānodrošina arhīva dokumentu pieejamība un izmantošana, kā arī to ierobežojumi saskaņā ar Arhīvu likumu. Tāpat sistēmai jānodrošina iespēja anonimizēt datus, ievērojot fizisko personu datu aizsardzības likumdošanu.

Vienotai informācijas sistēmai jābalstās uz atvērta koda risinājumiem, un tai ir jābūt pieejamai tīklā no jebkuras vietas, izmantojot lietotāja autentifikāciju.

Papildus jāizstrādā arī dažādi palīgriki, lai atvieglotu digitalizēto materiālu ievietošanu, satura atpazīšanu un metadatu pievienošanu.

Secinājumi

Finansējums digitalizācijai

Latvijas Nacionālajam arhīvam un Latvijas Valsts arhīvam tai skaitā ir ierobežoti resursi krājumu digitalizācijai. Arī VDK zinātniskās izpētes komisijas budžetā nav paredzēta nauda VDK rīcību atspoguļojošā dokumentārā mantojuma digitalizācijai, lai arī tas varētu veicināt komisijas pētniecības mērķu efektīvāku un kvalitatīvāku sasniegšanu. Viegli un ērti pieejami materiāli digitālā formā paver ceļu arī plašākiem un dziļākiem kvantitatīviem pētījumiem.

Digitalizācijas posmi

Materiālu digitalizācija ir jāsadalā loģiskos posmos tā, lai pēc katra no posmiem būtu reāls ieguvums lietotājiem. Pirmais posms ir materiālu ieskenēšana un attēlu iegūšana, lai būtu iespējama vienkārša šķirstīšana attālināti. Otrās digitalizācijas posms ir ieskenēto materiālu tekstu automātiska atpazīšana, ja tas vispār ir iespējams, ar vienkāršotu meklēšanu. Pēdējais posms sastāv no vairākiem apakšposmiem, kurus var izpildīt jebkurā secībā atkarībā no lietotāju prioritātēm — pilnīgu metadatu pievienošana par objektiem; automātiski atpazītā teksta pārbaudīšana vai teksta manuāla ievade; strukturēto datu pārvēršana datubāzes vai kādā citā formā, kas atbalsta vaicājumu izpildi; pilnvērtīga meklēšana pēc dažādiem parametriem.

Prioritāte ir kartotēku digitalizācija

Pašreiz arhīva apmeklētāji nevar nepastarpināti strādāt ar kartotēkām, kas pēc būtības ir fondu satura rādītāji, tāpēc īpaši svarīgi pētniecības darbā. Pat pēc digitalizācijas pirmā posma, kad kartotēka ir ieskenēta un sadalīta grupās (piemēram, pēc alfabēta), un to var vizuāli pāršķirstīt, ir būtisks uzlabojums arhīva lietotāju darbam. Otrai solis ir kartotēku satura atpazīšana un glabāšana strukturētā formātā. Priekšnosacījums, lai varētu veikt digitalizāciju, ir kartotēku aprakstīšana un sanumurēšana, kas vēl nav veikta visām kartotēkām.

Vienotā informācijas sistēma

Digitalizācijas un dažādu pētnieku darba procesā iegūtos un radītos metadatus un papilddatus nepieciešams saglabāt vienotā informācijas sistēmā. Tā kā īpaši svarīga ir dažādu personu un objektu savstarpējā saistība, tad sistēmas pamatā jābūt grafu datubāzei, kas ir piemērota sarežģītu attiecību glabāšanā un dažādu vaicājumu izpildei. Šai datubāzei jāvar pēc iespējas vieglāk mainīt struktūru (pievienot jaunu, pārveidot vecos datu laukus), jo darba gaitā atklāsies jaunas niansas un prasības, ko nevar paredzēt no paša sākuma.

Jebkura pētnieka darba starprezultāti un gala rezultāti satur lietderīgu informāciju un metadatus, kuri noteikti jā saglabā vienotajā informācijas sistēmā, lai citi lietotāji varētu atkārtoti izmantot jau vienreiz ieguldīta darba rezultātus.

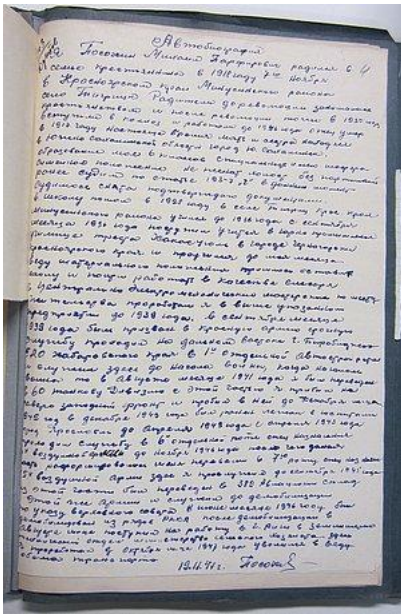
Īpaši jāuzsver digitālo materiālu ilgtermiņa saglabāšanas aspekts un ar to saistītās regulārās uzturēšanas izmaksas. Ja netiek atvēlēti līdzekļi šīs problēmas risināšanai, tad tas ir laika jautājums, kad tiks pazaudēti kādi digitālie materiāli.

1. pielikums

LNA Latvijas Valsts arhīvā esošie VDK, leTK izcelsmes fondi un ar tiem saistīto LKP dokumenti.[3]

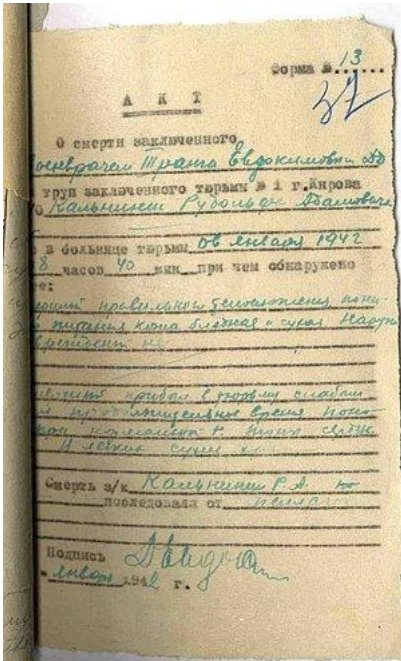
LVA-1821	PSRS Pārbaudes filtrācijas punktos un nometnēs ieslodzīto Latvijas iedzīvotāju personas lietas. 58451 lietas (642 961 lapas) + 91 000 kartītes.
LVA-1822	Latvijas PSR Valsts drošības ministrijas iznīcinātāju bataljoni (1944–1955). 3082 lietas (154 530 lapas)+ 44 000 kartītes
LVA-1824	Baltijas kara apgabala Kara prokuratūra (1941–1964) Reabilitēto personu uzraudzības lietas. 4012 lietas (253 316 lapas) + 8 000 kartītes
LVA-1825	Latvijas PSR Valsts drošības komitejas (VDK) dokumentu kolekcija (1940–1991). 502 lietas (100 008 lapas)
LVA-1846	Latvijā apglabāto vācu karagūstekņu kapu grāmatas (1945–1950). 1. apraksts. Lietvedība. 31 lieta (806 lapas)
LVA-1847	PSRS Iekšlietu ministrijas Iekšlietu karaspēka 5. strēlnieku divīzija (1944–1956). 1. apraksts. Lietvedība un pavēles. 16 lietas (2 528 lapas)
LVA-1894	1949. gada 25. martā no Latvijas izsūtīto iedzīvotāju personas lietas. (1949–1992). 13 429 lietas (1 102 550 lapas)
LVA-1986	Latvijas PSR Valsts drošības komitejas par sevišķi bīstamiem pretvalstiskiem noziegumiem apsūdzēto personu krimināllietas. 52 924 lietas (13 227 520 lapas)
LVA-1986	Latvijas represēto iedzīvotāju kartotēka (satur LVA-1894, LVA-1986, LVA-1987, LVA-1994 fondu kartiņas) Aptuveni 100 000 kartītes
LVA-1987	1941. gada 14. jūnijā no Latvijas izsūtīto iedzīvotāju personas lietas. 5165 lietas (444 190 lapas)
LVA-1994	1945.–1953. gadā no Latvijas izsūtīto vācu tautības iedzīvotāju, bezvalstnieku, reliģisko sektu dalībnieku un antisociālo elementu personas lietas. 1239 lietas (58 233 lapas)
LVA-PA-15500; LVA-PA-101 (102, 106, 107, 108, 109, 110, 111 utt.)	LKP CK Rīgas, Daugavpils, Jelgavas, Liepājas, Rēzeknes, Ventpils pilsētu partijas komiteju, LKJS fondu kartotēka (1944–1990) Alfabētiskā uzskaites kartotēka, kas aptver vadošos LKP un LKJS kadrus (LKP CK nomenklatūras kadri (biedri, kandidāti, LKJS biedri), Rīgas pilsētas, 8 Rīgas rajonu, 7 republikas pilsētu, 26 rajonu, 60 rajonu, 19 aprinču LKP un LKJS vadošie kadri, pavisam 121 LKP un ap 120 LKJS dažādi fondu nosaukumi) 237 000 kartītes
LVA-PA-106; LVA-PA-109	LKP Rīgas pilsētas Kirova rajona komiteja un LKP Rīgas pilsētas Proletāriešu rajonu komiteja. 82 000 kartītes + citi dokumenti
LVA-PA-15500	LKP CK nomenklatūras darbinieku personas lietu kolekcija (1940–1983). 13 991 lietas (1 280 510 lapas)

2. pielikums



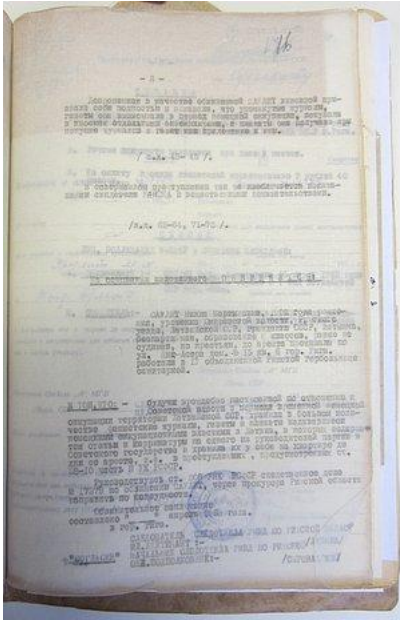
Autobiogrāfija, kuras saturs nav pilnībā izlasāms lapas iešuvuma dēļ.[4]

3. pielikums



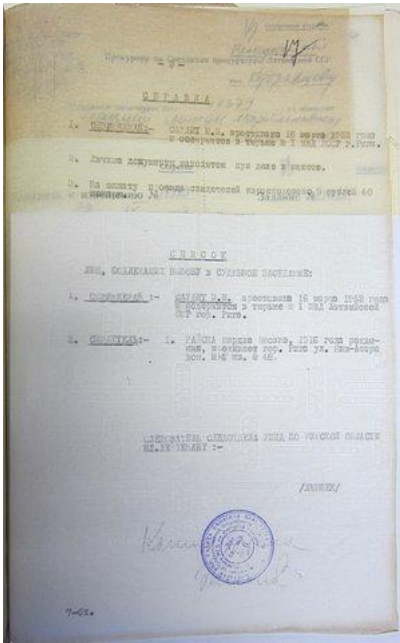
Akts par ieslodzītā nāvi, kura saturs nav pilnībā izlasāms lapas iešuvuma dēļ.[5]

4. pielikums



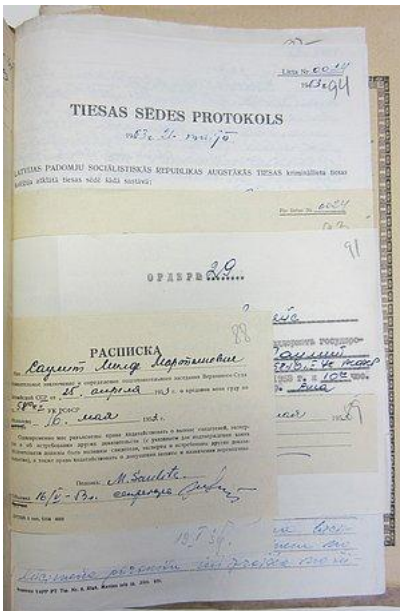
Puscaurspīdīgs papīrs.[6]

5. pielikums



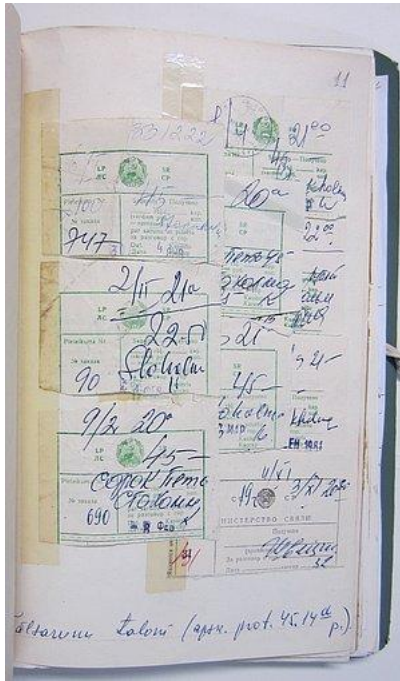
Puscaurspīdīgs papīrs, zem kura daļēji palikta balta papīra lapa.[7]

6. pielikums



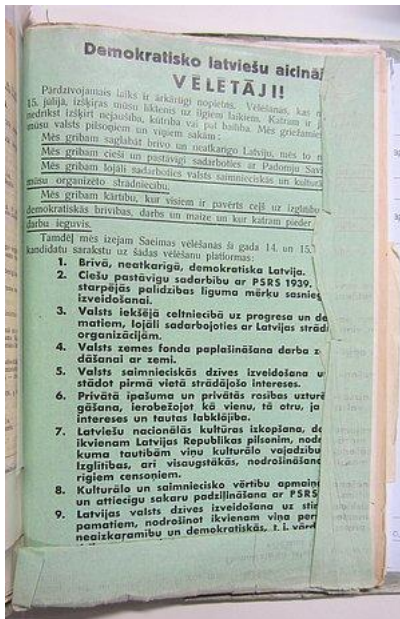
Lapu "vēdekļis" — mazākas lapas ir sašūtas, pakāpeniski tās nobīdot.[8]

7. pielikums



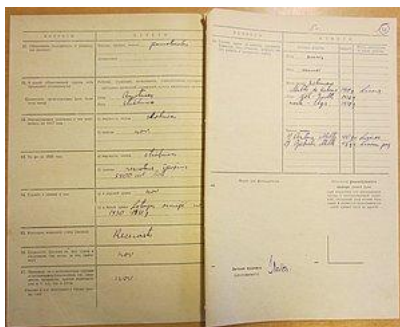
Tālsarunu taloni, kas pielīmēti ar līmlentu un daļēji aizklāj viens otru.[9]

8. pielikums



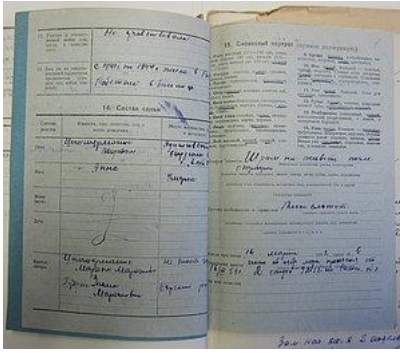
Aicinājums vēletājiem ar aizloctām malām.[10]

9. pielikums



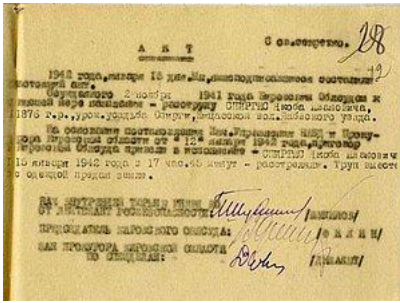
Anketa.[11]

10. pielikums



Anketa.[12]

11. pielikums



Izplūdis mašīnraksts, kas veikts caur kopppapīru.[13]

12. pielikums



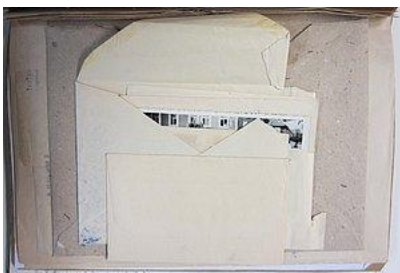
Pase.[14]

13. pielikums



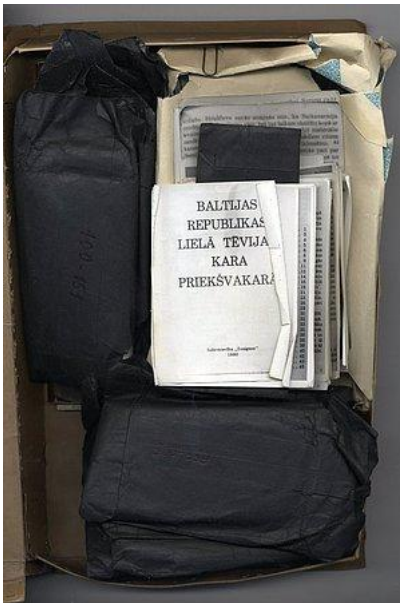
Aploksne.[15]

14. pielikums



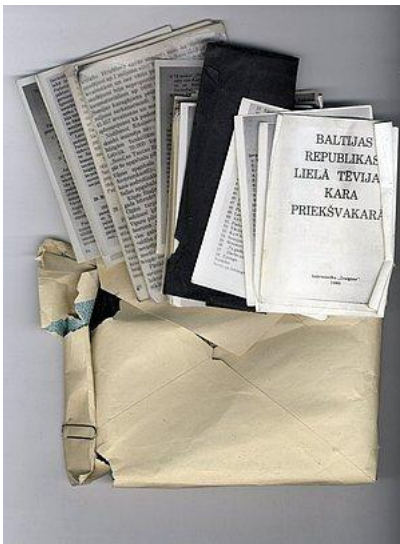
Atvērta aploksne.[16]

15. pielikums



Lietisko pierādījumu kaste.[17]

16. pielikums



Aploksne ar fotogrāfijām no lietisko pierādījumu kastes.[18]

17. pielikums



Kinolente.[19]

18. pielikums



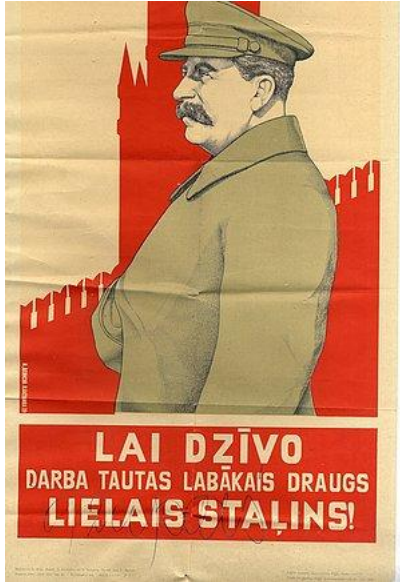
Filmiju negatīvi.[20]

19. pielikums



Fotogrāfija.[21]

20. pielikums



Lielformāta plakāts.[22]

Digitization of the documentary heritage reflecting the actions of the LSSR KGB

The largest volume of the documentary heritage reflecting the actions of the LSSR KGB is stored in the State Archive of Latvia of the Nation Archive of Latvia. Its collections contain more than 150 thousand related files and more than 1 million index cards. Digitized materials are much more accessible to researchers and other interested persons because they can be accessed online from any computer. Since there is no necessity to visit an archive in person, this saves both time and money. Moreover, this provides the possibility to access the materials which would otherwise not be accessible. The digitisation of materials helps to save the originals from damage or even permanent loss from depreciation and time decay caused by the disappearance of ink or the disintegration of paper. Digitisation should result in a reliable form of digital reproduction, which duplicates as fully as is possible the original object both in content and appearance. It is also important to retain the sequence and place of objects and their parts. The more accurately the original is reproduced, the more likely it will suit for research involving even specific analysis of the object. Digital archive files are intended for long-term preservation as well as for ensuring a second copy of each document. They are usually uncompressed and require little or no editing after digitizing. Access files (images) are derived from archive files and are intended for publication on the web and for viewing on the screen.

Index card files are important because they serve as tables of contents. When digitizing materials, the digitization of index cards should be prioritized. Scanning index cards is also easier to do; this can be executed in several steps.

The first step is the scanning of index cards, enabling researchers to look at card images in digital form. The second step is to recognize card content while storing it in a structured format. This allows researchers to undertake quantitative research using different cross-sections. Such research is not possible without the computer analysis of digitized materials.

By comparison, digitization of bound files is far more complicated. The following issues make it difficult: semi-transparent paper; glued paper sheets; folded larger pages; booklets; photos; three dimensional objects; one-page content that partially continues onto next page; wrapped sheet edges partially obscuring the content; and huge documents and posters.

Data and metadata created during the digitization of materials and the work of various researchers should be saved into one unified graph database, so all results can be reused in further research.

[1] *No NKVD līdz KGB. Politiskās prāvas Latvijā: 1940–1986. Noziegumos pret padomju valsti apsūdzēto Latvijas iedzīvotāju rādītājs.* Rīga: Latvijas vēstures institūta apgāds, 1999.

[2] Yuzhikov, Vladimir. *Restoration of defocused and blurred images.* Pieejams: yuzhikov.com/articles/BlurredImagesRestoration1.htm

[3] LVA esošo VDK, leTK izcelsmes fondu un ar tiem saistīto LKP dokumentu digitalizācijas izmaksas.

[4] LVA, 1822. f., 2. apr., 296. l., 4. lp.

[5] LVA, 1986. f., 1. apr., 41319. u.l., 37. lp.

[6] LVA, 1986. f., 1. apr., 21267 u.l., 16. lp.

[7] LVA, 1986. f., 1. apr., 21267 u.l., 17. lp.

[8] LVA, 1986. f., 1. apr., 21267. u.l., 88.–93. lp.

[9] LVA, 1986. f., 2. apr., P10894. l., 12. sēj., 11. lp.

[10] LVA, 1986. f., 2. apr., P5409. l., 65. lp.

[11] LVA, 1987. f., 1. apr., P16122. l., 3.–4. lp.

[12] LVA, 1986. f., 1. apr., 21267. u.l., 6.–7. lp.

[13] LVA, 1986. f., 1. apr., 44743. u.l., 12. lp.

[14] LVA, 1986. f., 1. apr., 44094. u.l., 15. lp.

[15] LVA, 1986. f., 2. apr., 5063. u.l., 31. lp.

[16] LVA, 1986. f., 2. apr., 5063. u.l., 31. lp.

[17] LVA, 1986. f., 1. apr., 45322. l., 14. sēj.

[18] LVA, 1986. f., 1. apr., 45322. l., 14. sēj.

[19] LVA, 1986. f., 1. apr., 45163. l.

[20] LVA, 1986. f., 1. apr., 10994. l., 11. sēj., 24. lp.

[21] LVA, 1986. f., 2. apr., 5063. u.l., 33. lp.

[22] LVA, 1986. f., 2. apr., P9397. u.l., 24. lp.

Saistītās ziņas:

Gints Zelmenis: 20. gadsimta astoņdesmito gadu «nepolitiskie» noziegumi okupētajā Latvijā un Valsts drošības komiteja 28.11.2017

Publicēta informācija par Ivara Godmaņa sadarbības fakta ar VDK pazudušās tiesas lietas meklēšanu 27.11.2017

Aldis Daugavvanags: Čekas loma kontrabandas izmeklēšanā (1945-1985). Atsevišķu piemēru analīze. 25.11.2017

Līnards Muciņš: Nošaušana sevišķā kārtībā - PSRS noziegums pret ius cogens normām. Liepājas «Zilā brīnuma» piemērs. 21.11.2017

Tiesiskais skatpunkts: Latvijas Republikas informācijas atklātības kods un VDK dokumenti. 20.11.2017

Tweet

Patīk 0

Ieteikt

